

Хранилища

Много данных

CP3: Minimize explicit sharing of writable data

- Клички
- Считаем слова
- Пишем опечаточник
- Балобол
- MapReduce
- Распределяем вычисления

Клички

name.txt

или names.txt?

Patrick Lucy Ignacio Carlo Rudy Carly Christi Rachael Cletus Lesley Cheri Abraham Clement Gerald Aimee Janet Susie Eula

Первые буквы

```
cat name.txt | cut -c -1
```

или

```
cut -c -1 name.txt
```

```
P L I C R C C R C L C A C G A J S E W D
```

```
sort
```

```
A A C C C C C C D E G I J L L P R R S W
```

```
uniq -c
```

```
2 A
```

```
6 C
```

```
1 D
```

```
1 E
```

```
1 G
```

```
1 I
```

```
1 J
```

```
2 L
```

```
1 P
```

```
2 R
```

```
1 S
```

```
1 W
```

```
uniq -c | sort -g
```

```
1 D
```

```
1 E
```

```
1 G
```

```
1 I
```

```
1 J
```

```
1 P
```

```
1 S
```

```
1 W
```

```
2 A
```

```
2 L
```

```
2 R
```

```
6 C
```

Или?

```
sort -gr
```

6 C
2 R
2 L
2 A
1 W
1 S
1 P
1 J
1 I
1 G
1 E
1 D

Или?

`uniq -d`

A
C
L
R

Для 2

Ca
Ch
Cl

Для 4

Carl

Для 5 решено

Abrah
Aimee
Carlo
Carly
Cheri
Chris
Cleme
Cletu
Doria
Eula
Geral
Ignac
Janet
Lesle
Lucy
Patri
Racha
Rudy
Susie
Willi

`cat name.txt | cut -c -5 | sort | uniq -d`

Считаем слова

I would have gotten the promotion, but my attendance wasn't good enough. The book is in front of the table. Wednesday is hump day, but has anyone asked the camel if he's happy about it? A glittering gem is not enough. A purple pig and a green donkey flew a kite in the middle of the night and ended up sunburnt. Two seats were vacant. Joe made the sugar cookies; Susan decorated them. The quick brown fox jumps over the lazy dog. How was the math test? I'd rather be a bird than a fish.

- Как делить, что разделитель?

- Регистр учитывать или нет?
- Апострофы?
- Формы слов?

a a a a a about and and anyone asked attendance be bird book brown but but camel cookies day decorated dog donkey ended enough enough fish flew fox front gem glittering good gotten green happy has have hes how hump i id if in in is is it joe jumps kite lazy made math middle my night not of of over pig promotion purple quick rather seats sugar sunburnt susan table test than the the the the the the the the the them two up vacant was wasnt wednesday were would

- 10 the
- 6 a
- 3 is
- 2 of
- 2 in
- 2 enough
- 2 but
- 2 and

Пишем очепятчик

I would have gotten the promotion, but my attendance wasn't good enough.

wou oul uld hav ave got ott tte ten the pro rom omo mot oti tio ion but att tte ten end nda dan anc nce was asn snt goo ood eno nou oug ugh

en [do]
go [ot]
ot [it]
ou [gl]
te [nn]
tt [ee]

Балобол

A -> B, C, D

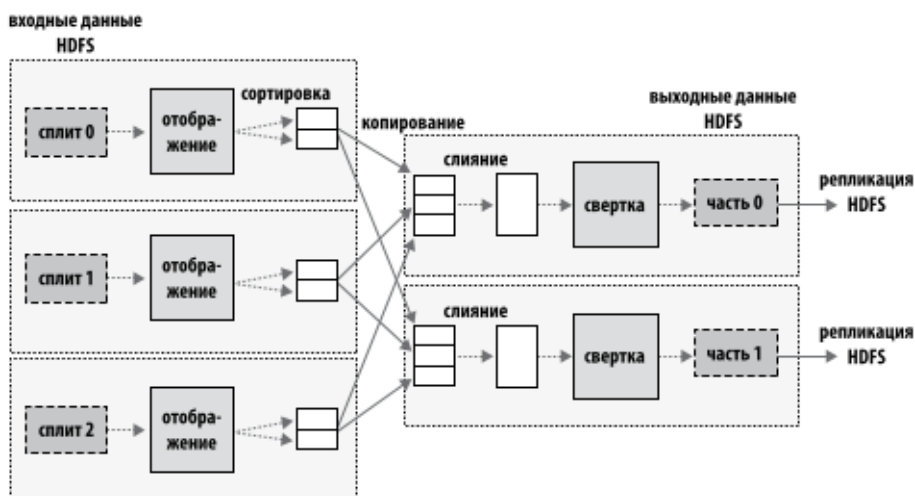
B -> D

B -> A

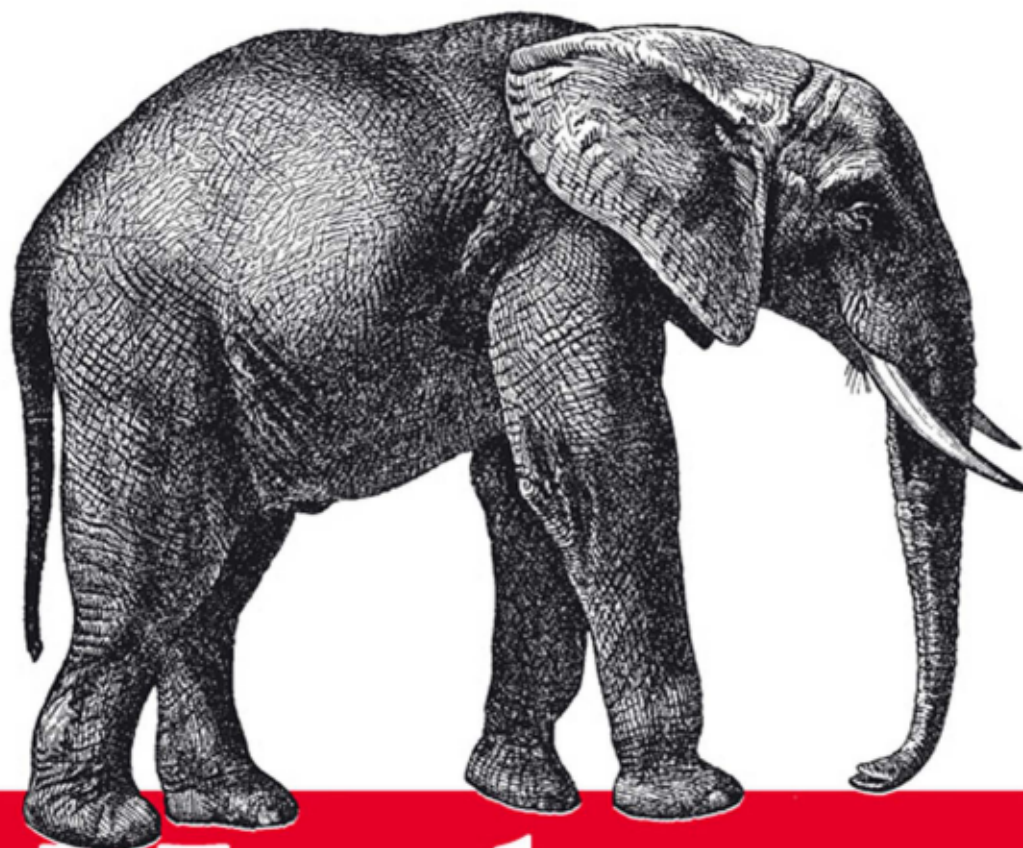
map -> sort -> reduce

Распределяем вычисления

Hadoop: The Definitive Guide 3rd Edition (978-1449311520)



Хранение и Анализ Интернет-масштаба



Hadoop

Подробное руководство

Том Уайт

O'REILLY®

 ПИТЕР®

Спасибо за внимание!